

# On the Optimality of Time Division for Broadcast Channels

Salman Beigi

*School of Mathematics, Institute for Research in Fundamental Sciences (IPM), Tehran, Iran*

*Department of Information Engineering, The Chinese University of Hong Kong, Hong Kong*

May 31, 2016

## Abstract

We provide a necessary and sufficient condition that under some technical assumption characterizes all two-receiver broadcast channels for which time division is optimal for transmission of private messages.

## 1 Introduction

A two-receiver discrete memoryless broadcast channel  $p(y, z|x)$  is a channel with one sender and two receivers. The goal of the sender is to send private messages to the receivers over multiple uses of the channel. The capacity region of the channel is the set of all rate pairs  $(R_1, R_2)$  such that private information with asymptotically vanishing error can be sent to the receivers at rate  $R_1$  and  $R_2$  respectively.

It is easy to verify that the capacity region of the broadcast channel  $p(y, z|x)$  depends only on the *marginal* channels  $p(y|x)$  and  $p(z|x)$  and not on the whole  $p(y, z|x)$ . So we may assume with no loss of generality that  $p(y, z|x) = p(y|x)p(z|x)$ . That is, we may think of a broadcast channel as two point-to-point channels with the same input sets.

There are some known inner and outer bounds for the capacity region of the broadcast channel [1], yet deriving a single letter formula for the capacity region is a long standing open problem. The best inner bound for the capacity region of the broadcast channel is due to Marton [2] and is described as follows. Let  $p_{UVWX}$  be an arbitrary distribution that induces the distribution  $p_{UVWXYZ}$ . Then any pair of non-negative numbers  $(R_1, R_2)$  satisfying

$$\begin{aligned} R_1 &\leq I(U, W; Y), \\ R_2 &\leq I(V, W; Z), \\ R_1 + R_2 &\leq \min \{I(W; Y), I(W; Z)\} + I(U; Y|W) + I(V; Z|W) - I(U; V|W), \end{aligned} \quad (1)$$

is an achievable rate pair.

The best outer bound on the capacity region of the broadcast channel is called the *UV* outer bound [3]. According to this outer bound for any achievable rate pair  $(R_1, R_2)$  there is a distribution  $p_{UVX}$  with the induced distribution  $p_{UVXYZ}$  such that

$$\begin{aligned} R_1 &\leq I(U; Y), \\ R_2 &\leq I(V; Z), \\ R_1 + R_2 &\leq \min \{I(U; Y) + I(V; Z|U), I(V; Z) + I(U; Y|V)\}. \end{aligned} \quad (2)$$

As mentioned above, the Marton inner bound (1) and the UV outer bound (2) do not match in general, and the capacity region of an arbitrary broadcast channel is not known.

A simple achievable rate region is derived by *time division*. Let  $C_1 = \max_{p_X} I(X; Y)$  be the capacity of the first channel  $p(y|x)$  and  $C_2 = \max_{p_X} I(X; Z)$  be the capacity of the second channel  $p(z|x)$ . By ignoring the second receiver, the sender can transmit information to the first receiver at the highest possible rate, namely  $C_1$ . Thus  $(R_1, R_2) = (C_1, 0)$  is achievable. Similarly  $(R_1, R_2) = (0, C_2)$  is in the capacity region. Moreover, the sender can use time sharing; she can send information to the first receiver in  $\alpha \in [0, 1]$  fraction of uses of the channel, and then send information to the second receiver in the remaining uses of the channel. Then the rate pair  $(R_1, R_2) = (\alpha C_1, (1 - \alpha)C_2)$ , for any  $\alpha \in [0, 1]$ , is achievable. More precisely, the whole set

$$\mathcal{R}_{\text{TD}} := \left\{ (R_1, R_2) : \frac{R_1}{C_1} + \frac{R_2}{C_2} \leq 1, R_1, R_2 \geq 0 \right\},$$

which we call the *time division* rate region, is in the capacity region.

The main result of this paper is a characterization of broadcast channels for which the time division rate region  $\mathcal{R}_{\text{TD}}$  is equal to the capacity region. Here is an informal statement of our main result.

**Theorem 1.** (Informal) Let  $p(y|x)$  and  $p(z|x)$  be two point-to-point channels with capacities  $C_1$  and  $C_2$  respectively. Suppose that  $C_1 \geq C_2$  and that the channels  $p(y|x), p(z|x)$  satisfy some technical assumptions. Then  $\mathcal{R}_{\text{TD}}$  is equal to the capacity region of the broadcast channel  $p(y, z|x) = p(y|x)p(z|x)$  if and only if either  $C_1 < C_2$  and

$$\frac{I(X; Y)}{C_1} \leq \frac{I(X; Z)}{C_2}, \quad \forall p_X, \quad (3)$$

or  $C_1 = C_2$  and the two channels are more capable comparable.

Recall that a channel  $p(y|x)$  is called *more capable* than  $p(z|x)$  if for all input distributions  $p_X$  we have

$$I(X; Y) \geq I(X; Z).$$

We say that two channels  $p(y|x)$  and  $p(z|x)$  are more capable comparable if either  $p(y|x)$  is more capable than  $p(z|x)$  or vice versa.

A partial characterization of *degraded* broadcast channels for which time division is optimal is provided in [4, Theorem 3] that is similar to our characterization.

To prove this theorem we use some known facts about the set of *capacity achieving distributions* of a point-to-point channel [5, Theorem 13.1.1]. For the convenience of the reader we also present the proofs of these facts.

## 2 Capacity achieving distributions

We use quite standard notations in this paper (see, e.g., [1]). Sets are denoted by calligraphic letters such as  $\mathcal{X}$ , and a distribution on such a set is specified by a subscript as in  $p_X$ . Discrete memoryless point-to-point channels are determined by a set of conditional distributions  $\{p_{Y|x}, x \in \mathcal{X}\}$ , on a set  $\mathcal{Y}$ . For ease of notation we denoted such a channel by  $p(y|x)$ . For  $\lambda \in [0, 1]$  we use the notation  $\bar{\lambda} := 1 - \lambda$ . To avoid confusions, when a mutual information  $I(X; Y)$  is computed with respect to a distribution  $p_{XY}$  we denote it by  $I(X; Y)_p$ .

Let  $p(y|x)$  be a discrete memoryless point-to-point channel with capacity  $C = \max_{p_X} I(X; Y)$ . For arbitrary distributions  $p_X$  and  $r_Y$  define

$$\psi(p_X, r_Y) = \sum_x p(x) D(p_{Y|x} \| r_Y) = D(p_{XY} \| p_X r_Y) = I(X; Y)_p + D(p_Y \| r_Y),$$

where  $D(\cdot \| \cdot)$  is the KL divergence, and  $p_{XY}$  is the induced distribution on the input and output of the channel with input distribution  $p_X$ . Then by the joint convexity of KL divergence and Sion's minimax theorem we have

$$\max_{p_X} \min_{r_Y} \psi(p_X, r_Y) = \min_{r_Y} \max_{p_X} \psi(p_X, r_Y).$$

Let us compute each side of the above equation. By the non-negativity of KL divergence we have

$$\max_{p_X} \min_{r_Y} \psi(p_X, r_Y) = \max_{p_X} \min_{r_Y} I(X; Y)_p + D(p_Y \| r_Y) = \max_{p_X} I(X; Y) = C.$$

On the other hand, by the linearity of  $\psi$  in  $p_X$  have

$$\min_{r_Y} \max_{p_X} \psi(p_X, r_Y) = \min_{r_Y} \max_{x_0} D(p_{Y|x_0} \| r_Y).$$

As a result,

$$\min_{r_Y} \max_{x_0} D(p_{Y|x_0} \| r_Y) = C. \quad (4)$$

Observe that the minimum in (4) is achieved. So let  $r_Y^*$  be some optimal distribution there, i.e.,  $r_Y^*$  is such that

$$\max_{x_0} D(p_{Y|x_0} \| r_Y^*) = \max_{p_X} I(X; Y)_p + D(p_Y \| r_Y^*) = C.$$

Then for every  $p_X$  we have

$$\psi(p_X, r_Y^*) = I(X; Y)_p + D(p_Y \| r_Y^*) \leq C. \quad (5)$$

Let  $\Pi$  be the set of capacity achieving distributions:

$$\Pi = \arg \max_{p_X} I(X; Y).$$

Then by (5) for every  $p_X \in \Pi$  we have  $D(p_Y \| r_Y^*) = 0$ , i.e.,  $p_Y = r_Y^*$ . This means that, for any capacity achieving distribution  $p_X \in \Pi$  its induced distribution on the output of the channel is fixed, i.e.,  $r_Y^* = p_Y$ . Indeed, the *optimal output distribution* of a channel is unique, and is the unique distribution  $r_Y^*$  that achieves the minimum in (4).

Let us define

$$\mathcal{K} := \{x : D(p_{Y|x} \| r_Y^*) = C\}. \quad (6)$$

Note that by the above discussion  $\mathcal{K}$  is non-empty. Let  $p_X$  be some distribution with  $\text{supp}(p_X) \subseteq \mathcal{K}$ , where  $\text{supp}(p_X) := \{x : p(x) > 0\}$ . Then we have

$$I(X; Y)_p + D(p_Y \| r_Y^*) = \sum_x p(x) D(p_{Y|x} \| r_Y^*) = \sum_{x \in \mathcal{K}} p(x) D(p_{Y|x} \| r_Y^*) = C.$$

This means that the inequality in (5) becomes an equality for all  $p_X$  with  $\text{supp}(p_X) \subseteq \mathcal{K}$ .

On the other hand, let  $p_X \in \Pi$  be some capacity achieving distribution. Then by the above discussion,  $p_Y = r_Y^*$ . Moreover, we have

$$\begin{aligned} C &= I(X; Y)_p = \sum_x p(x) D(p_{Y|x} \| r_Y^*) \\ &= \sum_{x \in \mathcal{K}} p(x) D(p_{Y|x} \| r_Y^*) + \sum_{x \notin \mathcal{K}} p(x) D(p_{Y|x} \| r_Y^*) \\ &= p(\mathcal{K})C + p(\mathcal{X} \setminus \mathcal{K}) \max_{x \notin \mathcal{K}} D(p_{Y|x} \| r_Y^*). \end{aligned}$$

As a result, we must have  $p(\mathcal{X} \setminus \mathcal{K}) = 0$ , i.e.,  $\text{supp}(p_X) \subseteq \mathcal{K}$ .

Finally, suppose that  $p_X$  is some distribution with  $\text{supp}(p_X) \subseteq \mathcal{K}$  and  $p_Y = r_Y^*$ . Then (5) becomes equality for  $p_X$  and since  $D(p_Y \| r_Y^*) = 0$ ,  $p_X$  is capacity achieving.

We summarize the above findings in the following proposition.

**Proposition 2.** *For any point-to-point channel  $p(y|x)$  there is a unique distribution  $r_Y^*$  such that for all capacity achieving distributions  $p_X \in \Pi$  we have  $p_Y = r_Y^*$ . Moreover, for any  $p_X \in \Pi$  we have  $\text{supp}(p_X) \subseteq \mathcal{K}$  where  $\mathcal{K}$  is defined in (6). Indeed, a given distribution  $p_X$  is capacity achieving if and only if  $\text{supp}(p_X) \subseteq \mathcal{K}$  and  $p_Y = r_Y^*$ . In particular  $\Pi$  is convex.*

The above proposition motivates the following definition. Define  $\mathcal{K}_0$  to be the union of the supports of capacity achieving distributions, i.e.,

$$\mathcal{K}' := \bigcap_{p_X \in \Pi} \text{supp}(p_X). \quad (7)$$

By the above proposition  $\mathcal{K}' \subseteq \mathcal{K}$ .

For a channel that has a capacity achieving distribution with full support (e.g., a channel for which the uniform distribution is capacity achieving) we have  $\mathcal{K}' = \mathcal{K} = \mathcal{X}$ . For example, this equality holds for binary symmetric and binary erasure channels. Later we will see an example of a channel for which the inclusion  $\mathcal{K}' \subseteq \mathcal{K}$  is strict.

**Proposition 3.** *There exists  $r_X \in \Pi$  such that  $\text{supp}(r_X) = \mathcal{K}'$ . Moreover, for any  $p_X$  with  $\text{supp}(p_X) \subseteq \mathcal{K}'$  we have*

$$I(X; Y)_p + D(p_Y \| r_Y^*) = \sum_{x \in \mathcal{K}'} p(x) D(p_{Y|x} \| r_Y^*) = C.$$

*Proof.* The existence of  $r_X \in \Pi$  with  $\text{supp}(r_X) = \mathcal{K}'$  follows from the definition of  $\mathcal{K}'$  and the convexity of  $\Pi$  established in Proposition 2. The second claim follows from  $\mathcal{K}' \subseteq \mathcal{K}$ .  $\square$

### 3 Proof of the main result

Let  $p(y|x)$  and  $p(z|x)$  be two channels with capacities  $C_1$  and  $C_2$  respectively. Let  $r_Y^*$  and  $s_Z^*$  be the optimal output distributions of the channels (as defined in the previous section) respectively. Also let  $\Pi_1$  and  $\Pi_2$  be their associated sets of capacity achieving distributions respectively. Finally let  $\mathcal{K}_1, \mathcal{K}'_1$  and  $\mathcal{K}_2, \mathcal{K}'_2$  be their associated subsets of  $\mathcal{X}$  defined by (6) and (7). Here is the formal statement of our main result.

**Theorem 4.** *Suppose that  $\mathcal{K}'_1 = \mathcal{K}'_2 = \mathcal{X}$  and  $C_1 \geq C_2$ . Then the time division rate region  $\mathcal{R}_{\text{TD}}$  is the capacity region of the broadcast channel  $p(y, z|x) = p(y|x)p(z|x)$  if and only if either  $C_1 < C_2$  and*

$$\frac{I(X; Y)}{C_1} \leq \frac{I(X; Z)}{C_2}, \quad \forall p_X, \quad (8)$$

*or  $C_1 = C_2$  and the two channels are more capable comparable.*

The rest of this section is devoted to the proof of this theorem.

( $\Rightarrow$ ) First suppose that the time division region is the capacity region. That is, for any achievable rate pair  $(R_1, R_2)$  we have

$$\frac{R_1}{C_1} + \frac{R_2}{C_2} \leq 1. \quad (9)$$

Let  $r_{XU}$  and  $s_{XV}$  be arbitrary distributions. Define  $p_{QW\tilde{U}\tilde{V}X}$  by

$$p(Q=0) = \lambda, \quad p(Q=1) = \bar{\lambda} = 1 - \lambda, \quad (10)$$

and according to the following table:

	$W$	$\tilde{U}$	$\tilde{V}$
$Q=0$	$U$	$X$	Const.
$Q=1$	$V$	Const.	$X$

This table should be understood as follows. First we have  $\mathcal{Q} = \{0, 1\}$  and the marginal distribution  $p_Q$  is given by (10). Second, we have  $\mathcal{W} = \mathcal{U} \cup \mathcal{V}$ ,  $\tilde{\mathcal{U}} = \mathcal{X} \cup \{u^*\}$  and  $\tilde{\mathcal{V}} = \mathcal{X} \cup \{v^*\}$  for two distinguished elements  $u^*, v^*$ . Third, the conditional distribution  $p(w, \tilde{u}, \tilde{v}, x|q)$  is given by

$$p(w, \tilde{u}, \tilde{v}, x|Q=0) = \begin{cases} r(X=x, U=w), & \tilde{u}=x, \tilde{v}=v^*, w \in \mathcal{U}, \\ 0, & \text{otherwise,} \end{cases}$$

and

$$p(w, \tilde{u}, \tilde{v}, x|Q=1) = \begin{cases} s(X=x, V=w), & \tilde{v}=x, \tilde{u}=u^*, w \in \mathcal{V}, \\ 0, & \text{otherwise.} \end{cases}$$

Now let  $\tilde{W} = (Q, W)$  and consider the distribution  $p_{\tilde{W}\tilde{U}\tilde{V}X Y Z}$  induced by the channel. Observe that  $I(\tilde{U}; \tilde{V}|\tilde{W}) = I(\tilde{U}; \tilde{V}|Q, W) = 0$ . Then by Marton's coding theorem (1) the rate pair  $(R_1, R_2)$  given by

$$\begin{cases} R_2 = I(\tilde{V}\tilde{W}; Z) = I(\tilde{W}; Z) + I(\tilde{V}; Z|\tilde{W}), \\ R_1 + R_2 = \min \{I(\tilde{W}; Y), I(\tilde{W}; Z)\} + I(\tilde{U}; Y|\tilde{W}) + I(\tilde{V}; Z|\tilde{W}), \end{cases}$$

is achievable. Therefore, by our assumption we must have

$$\frac{R_1}{C_1} + \frac{R_2}{C_2} = \frac{R_1 + R_2}{C_1} + \left(\frac{1}{C_2} - \frac{1}{C_1}\right)R_2 \leq 1,$$

and then

$$\frac{1}{C_1} \min \{I(\tilde{W}; Y), I(\tilde{W}; Z)\} + \left(\frac{1}{C_2} - \frac{1}{C_1}\right)I(\tilde{W}; Z) + \frac{1}{C_1}I(\tilde{U}; Y|\tilde{W}) + \frac{1}{C_2}I(\tilde{V}; Z|\tilde{W}) \leq 1. \quad (11)$$

Let us compute individual terms in the above equation. We have

$$I(\tilde{W}; Y) = I(Q, W; Y) = I(Q; Y) + I(W; Y|Q) = I(Q; Y) + \lambda I(U; Y)_r + \bar{\lambda} I(V; Y)_s. \quad (12)$$

We similarly have

$$I(\widetilde{W}; Z) = I(Q; Z) + \lambda I(U; Z)_r + \bar{\lambda} I(V; Z)_s. \quad (13)$$

Moreover, observe that

$$I(\widetilde{U}; Y|\widetilde{W}) = \lambda I(X; Y|U)_r,$$

and

$$I(\widetilde{V}; Z|\widetilde{W}) = \bar{\lambda} I(X; Z|V)_s.$$

Putting the above two equations in (11) we find that

$$\frac{1}{C_1} \min \{I(\widetilde{W}; Y), I(\widetilde{W}; Z)\} + \left(\frac{1}{C_2} - \frac{1}{C_1}\right) I(\widetilde{W}; Z) + \frac{\lambda}{C_1} I(X; Y|U)_r + \frac{\bar{\lambda}}{C_2} I(X; Z|V)_s \leq 1. \quad (14)$$

We can now consider two cases: either  $I(\widetilde{W}; Y) \geq I(\widetilde{W}; Z)$  or  $I(\widetilde{W}; Y) < I(\widetilde{W}; Z)$ . Then using  $C_1 \geq C_2$ , equations (12) and (13), and ignoring some non-negative terms in (14) we find that

$$\begin{cases} \frac{1}{C_1} \left( \lambda I(U; Y)_r + \bar{\lambda} I(V; Y)_s \right) + \left( \frac{1}{C_2} - \frac{1}{C_1} \right) \bar{\lambda} I(V; Z)_s + \frac{\lambda}{C_1} I(X; Y|U)_r + \frac{\bar{\lambda}}{C_2} I(X; Z|V)_s \leq 1, \\ \text{or} \\ \frac{1}{C_2} \left( \lambda I(U; Z)_r + \bar{\lambda} I(V; Z)_s \right) + \frac{\lambda}{C_1} I(X; Y|U)_r + \frac{\bar{\lambda}}{C_2} I(X; Z|V)_s \leq 1. \end{cases}$$

Now suppose that we chose  $r_{XU}$  and  $s_{XV}$  such that  $r_X \in \Pi_1$  and  $s_X \in \Pi_2$ . Then using  $C_1 = I(X; Y)_r = I(UX; Y)_r = I(U; Y)_r + I(X; Y|U)_r$  and  $C_2 = I(X; Z)_s = I(VX; Z)_s = I(U; Z)_s + I(X; Z|V)_s$ , by a simple algebra we arrive at

$$\begin{cases} I(V; Y)_s \leq I(V; Z)_s, \\ \text{or} \\ \frac{1}{C_2} I(U; Z)_r \leq \frac{1}{C_1} I(U; Y)_r. \end{cases}$$

Observe that the first inequality here depends only on  $s_{XV}$  and the second one is solely in terms of  $r_{XU}$ . Then either the first one holds for every valid choice of  $s_{XV}$  or the second one holds for every valid choice of  $r_{XU}$ . This means that either

$$I(V; Y) \leq I(V; Z), \quad \forall s_{VX} \text{ s.t. } s_X \in \Pi_2, \quad (15)$$

or

$$\frac{1}{C_2} I(U; Z) \leq \frac{1}{C_1} I(U; Y), \quad \forall r_{UX} \text{ s.t. } r_X \in \Pi_1. \quad (16)$$

Let us suppose that (15) holds. Fix  $s_X \in \Pi_2$  to be a capacity achieving distribution for  $p(z|x)$  with  $\text{supp}(s_X) = \mathcal{K}'_2 = \mathcal{X}$  whose existence is guaranteed by Proposition 3. Let  $p_X$  be an arbitrary distribution. Define  $s_{VX}$  as follows. Let  $\mathcal{V} = \{0, 1\}$  and define  $s(V=0) = \epsilon$  and  $s(V=1) = 1 - \epsilon$ . Also let

$$s(x|V=0) = p(x), \quad s(x|V=1) = \frac{1}{1-\epsilon} s(x) - \frac{\epsilon}{1-\epsilon} p(x).$$

Observe that  $\text{supp}(p_X) \subseteq \text{supp}(s_X) = \mathcal{X}$ , so for sufficiently small  $\epsilon > 0$ , both  $s(x|V=0)$  and  $s(x|V=1)$  are valid distributions. Then we obtain a distribution  $s_{VX}$  on  $\{0, 1\} \times \mathcal{X}$  whose marginal on  $\mathcal{X}$  is the distribution  $s_X \in \Pi_2$  we started with. Since we assumed that (15) holds, for any sufficiently small  $\epsilon > 0$  we have  $I(V; Y) \leq I(V; Z)$ . Now a simple computation verifies that  $I(V; Y) = \epsilon D(p_Y \| s_Y) + \Theta(\epsilon^2)$  and  $I(V; Z) = \epsilon D(p_Z \| s_Z) + \Theta(\epsilon^2)$ . Therefore, we have

$$D(p_Y \| s_Y) \leq D(p_Z \| s_Z^*), \quad \forall p_X. \quad (17)$$

Starting from (16) and following similar arguments we find that

$$\frac{1}{C_2} D(p_Z \| r_Z) \leq \frac{1}{C_1} D(p_Y \| r_Y^*), \quad \forall p_X, \quad (18)$$

where  $r_X \in \Pi_2$  is a capacity achieving distribution for  $p(y|x)$  with  $\text{supp}(r_X) = \mathcal{K}'_1 = \mathcal{X}$ . Then either (17) or (18) is satisfied.

Let us in (17) and (18) restrict ourself to  $p_X$  of the form  $p(x) = \delta_{x,x_0}$ , where  $\delta_{x,x_0}$  denotes the Kronecker delta function and  $x_0 \in \mathcal{X}$  is arbitrary. Then either

$$D(p_{Y|x_0} \| s_Y) \leq D(p_{Z|x_0} \| s_Z^*) = C_2, \quad \forall x_0, \quad (19)$$

or

$$\frac{1}{C_2} D(p_{Z|x_0} \| r_Z) \leq \frac{1}{C_1} D(p_{Y|x_0} \| r_Y^*) = 1, \quad \forall x_0, \quad (20)$$

holds.

If (17) and then (19) hold, we have  $\max_{x_0} D(p_{Z|x_0} \| s_Y) \leq C_2$ , which using (4) gives  $C_1 \leq C_2$ . Then by our assumption  $C_1 \geq C_2$  we arrive at  $C_1 = C_2$ . Therefore, (17) does not hold if  $C_1 > C_2$ . Moreover, if  $C_1 = C_2$ , (17) and (18) are symmetric. So in both cases, with no loss of generality we may assume that (18) and then (20) are satisfied.

We note that (20) implies that  $\max_{x_0} D(p_{Z|x_0} \| r_Z) \leq C_2$ . Thus  $r_Z$  is an optimal output distribution for  $p(z|x)$ , and by its uniqueness  $r_Z = s_Z^*$ . Then (18) reduces to

$$\frac{1}{C_2} D(p_Z \| s_Z^*) \leq \frac{1}{C_1} D(p_Y \| r_Y^*), \quad \forall p_X.$$

On the other hand, by Proposition 3 we have  $D(p_Y \| r_Y^*) = C_1 - I(X; Y)_p$  and  $D(p_Z \| s_Z^*) = C_2 - I(X; Z)_p$ . Using these in the above inequality gives (8).

( $\Leftarrow$ ) We now prove the converse. We assume that either  $C_1 > C_2$  and (8) holds, or  $C_1 = C_2$  and the two channels are more-capable comparable. In the latter case, by symmetry with no loss of generality we assume that  $p(y|x)$  is more capable than  $p(z|x)$ . Then in both cases (8) holds. Assuming this we show that time division is optimal for the broadcast channel  $p(y|x)p(z|x)$ .

Let  $(R_1, R_2)$  be an achievable rate pair. By the  $UV$  outer bound (2), there exists  $p_{UVX}$  such that

$$\begin{aligned} R_2 &\leq I(V; Z), \\ R_1 + R_2 &\leq I(V; Z) + I(U; Y|V) \leq I(V; Z) + I(X; Y|V). \end{aligned}$$

Then using the fact that  $C_1 \geq C_2$  we find that

$$\begin{aligned} \frac{R_1}{C_1} + \frac{R_2}{C_2} &\leq \frac{I(X; Y|V)}{C_1} + \frac{I(V; Z)}{C_2} \\ &\leq \frac{I(X; Z|V)}{C_2} + \frac{I(V; Z)}{C_2} \\ &\leq \frac{I(X; Z)}{C_2} \\ &\leq 1, \end{aligned}$$

where in the second line we use (8). We are done.

## 4 Example

In the statement of Theorem 4 we assume that  $\mathcal{K}'_1 = \mathcal{K}'_2 = \mathcal{X}$ . This may seem an unnecessary technical assumption that is forced by our proof method. Here we give an example to illustrate that Theorem 4 does not hold without it.

Let  $\mathcal{A}, \mathcal{B}$  be two finite disjoint sets with  $|\mathcal{A}| \geq |\mathcal{B}| \geq 2$ . Let  $\mathcal{X} = \mathcal{A} \cup \mathcal{B}$ ,  $\mathcal{Y} = \mathcal{A}$  and  $\mathcal{Z} = \mathcal{B}$ . Define channels  $p(y|x)$  and  $p(z|x)$  by

$$p(y|x) = \begin{cases} \delta_{x,y} & x \in \mathcal{A}, \\ \frac{1}{|\mathcal{B}|} & x \in \mathcal{B}, \end{cases} \quad p(z|x) = \begin{cases} \frac{1}{|\mathcal{B}|} & x \in \mathcal{A}, \\ \delta_{z,x} & x \in \mathcal{B}, \end{cases}$$

where  $\delta$  denotes the Kronecker delta function.

The uniform distribution on the subset  $\mathcal{A} \subset \mathcal{X}$  is the unique capacity achieving distribution of  $p(y|x)$ , and its optimal output distribution is the uniform distribution on  $\mathcal{Y}$ . Then  $\mathcal{K}'_1 = \mathcal{A}$ . Likewise the unique

capacity achieving distribution of  $p(z|x)$  is the uniform distribution on  $\mathcal{B} \subset \mathcal{X}$  and we have  $\mathcal{K}'_2 = \mathcal{B}$ . Moreover,  $C_1 = \log |\mathcal{A}| \geq \log |\mathcal{B}| = C_2$ . Also note that (8) does not hold, nor the two channels are more-capable comparable. Indeed, if  $p_X$  is a non-trivial distribution supported only on  $\mathcal{A}$ , then  $I(X; Y)_p > 0$  while  $I(X; Z)_p = 0$ . Similarly if  $p_X$  is non-trivial and supported only on  $\mathcal{B}$ , then  $I(X; Y)_p = 0$  while  $I(X; Z)_p > 0$ . Nevertheless, we show in the following that time division is optimal for the broadcast channel  $p(y|x)p(z|x)$ .

Let  $(R_1, R_2)$  be an achievable rate pair. Again by the  $UV$  outer bound (2), there exists  $p_{UVX}$  such that

$$\begin{aligned} R_2 &\leq I(V; Z), \\ R_1 + R_2 &\leq I(V; Z) + I(U; Y|V) \leq I(V; Z) + I(X; Y|V). \end{aligned}$$

Then using the fact that  $C_1 \geq C_2$  we find that

$$\begin{aligned} \frac{R_1}{C_1} + \frac{R_2}{C_2} &\leq \frac{I(X; Y|V)}{C_1} + \frac{I(V; Z)}{C_2} \\ &\leq \frac{I(X; Y)}{C_1} + \frac{I(V; Z)}{C_2}, \end{aligned} \tag{21}$$

where in the second line we use the fact that  $V - X - Y$  forms a Markov chain.

Let  $Q$  be a binary random variable that equals 0 if  $X \in \mathcal{A}$  and equals 1 if  $X \in \mathcal{B}$ . Then we have

$$\begin{aligned} I(X; Y) &= I(XQ; Y) \\ &= H(Y) - p(Q=0)H(Y|X, Q=0) - p(Q=1)H(Y|X, Q=1) \\ &= H(Y) - p(Q=1)\log |\mathcal{A}| \\ &\leq p(Q=0)C_1. \end{aligned}$$

We similarly have  $I(X; Z) \leq p(Q=1)C_2$ . Putting these in (21) we find that  $R_1/C_1 + R_2/C_2 \leq 1$ . Therefore, time division is optimal for  $p(y|x)p(z|x)$ .

**Acknowledgements.** The author is thankful to Chandra Nair for several fruitful discussions about broadcast channels. The author was supported in part by Institute of Network Coding of CUHK and by GRF grants 2150829 and 2150785.

## References

- [1] A. El Gamal and Y.-H. Kim, Network information theory, Cambridge university press (2011).
- [2] K. Marton, A coding theorem for the discrete memoryless broadcast channel, IEEE Trans. Info. Theory **25**, 306-311 (1979).
- [3] C. Nair and A. El Gamal, An outer bound to the capacity region of the broadcast channel, IEEE Trans. Info. Theory **53**, 350-355 (2007).
- [4] A. Gohari, A. El Gamal and V. Anantharam, On Marton's Inner Bound for the General Broadcast Channel, IEEE Trans. Info. Theory **60**, 3748-376 (2014).
- [5] T. M. Cover and J. A. Thomas, Elements of Information Theory, John Wiley & Sons (2012).